Euphytica (2008) 161:61–69
DOI 10.1007/s10681-007-9509-0

QTL × genetic background interaction: predicting inbred progeny value

Jean-Luc Jannink

Received: 20 April 2007 / Accepted: 11 July 2007 / Published online: 16 August 2007

Abstract Failures of the additive infinitesimal model continue to provide incentive to study other modes of gene action, in particular, epistasis. Epistasis can be modeled as a QTL by genetic background interaction. Association mapping models lend themselves to fitting such an interaction because they often include both main marker and genetic background factors. In this study, I review a model that fits the QTL by background interaction as an added random effect in the now standard mixed model framework of association analyses. The model is applied to four-generation pedigrees where the objective is to predict the genotypic values of fourth-generation individuals that have not been phenotyped. In particular, I look at how well epistatic effects are estimated under two levels of inbreeding. Interaction detection power was 8% and 65% for pedigrees of 240 randomly mated individuals when the interaction generated 6% and 20% of the phenotypic variance, respectively. Power increased to 21% and 94% for these conditions when evaluated individuals were inbred by selfing four times. The interaction variance was estimated in an unbiased way under both levels of inbreeding, but its mean squared error was reduced by 40% to 70% when estimated in inbred relative to randomly mated individuals. The performance of the epistatic model was also enhanced relative to the additive model by inbreeding. These results are promising for the application of the model to typically self-pollinating crops such as wheat and soybean.

Keywords Association mapping · Epistasis · Genetic background · Inbreeding · Pedigree

Abbreviations
IBD Identity by descent
QTL Quantitative trait locus/loci

Introduction

The additive infinitesimal model has served remarkably well, both for the development of quantitative genetics and for the prediction of resemblance between relatives and short-term genetic gain. Nevertheless, failures of this model continue to provide incentive to study other modes of gene action, in particular, epistasis. A parameter that additive models fail to predict well is the genetic variance, especially for mid- to long-term selection in which some level of inbreeding is occurring (Fig. 1). The effect of epistasis on this parameter has been studied theoretically (Cheverud and Routman 1996; Goodnight 1987; Goodnight 2004; Jannink 2003), and

J.-L. Jannink (✉)
US Plant, Soil & Nutrition Laboratory, USDA-ARS,
Tower Road, Ithaca, NY 14853-2901, USA
e-mail: JeanLuc.Jannink@ars.usda.gov

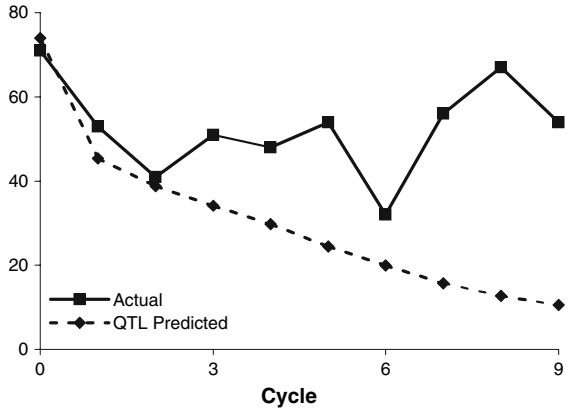

Fig. 1 Observed and predicted response of the genotypic variance to selection for high oil percentage of genotypic variance in oil content. Observed response was documented in Frey and Holland (1999). Predicted response derived from simulation using Qu-Gene (Podlich and Cooper 1998) and assuming 70 QTL with additive effects equal to 1/10 of those identified in Kianian et al. (1999) and with initial frequencies sampled from a uniform distribution

observations consistent with its predicted impact have been observed after founder events (Bryant and Meffert 1993; Cheverud et al. 1999) and during selection (Fig. 1; Carlborg et al. 2006).

A further incentive to study epistasis is that biotechnologies are giving us unprecedented ability to determine what alleles are present at loci of interest. To take advantage of this ability, we need to be able to predict what phenotypes are conferred by the multi-locus genotypes that we can assemble. Then the process of crop improvement will be turned on its head: rather than selecting phenotypes of interest in order to obtain desired genotypes, we will select target genotypes in order to obtain desired phenotypes. This enduringly elusive "improvement by design" clearly requires greater ability to predict phenotype from multi-locus genotype than we currently have.

One approach to studying epistasis is to examine the effect of alleles at a QTL as modulated by the genetic background in which they are placed. This approach is appealing for two reasons. First, it is simple. Allowing for the genetic background to serve as the interactor removes the need to know the identity of the specific interacting loci or how many such loci there are. It also decreases the importance of the mode of interaction (additive by additive, additive by dominance, additive by additive by additive, etc.). Finally, in the search for loci that act epistatically, it allows for a parsimonius one-dimensional search strategy (Boer et al. 2002; Jannink and Jansen 2001).

This interest in epistasis comes at a time when association genetic studies are demonstrating their ability to identify marker polymorphisms correlated with the phenotype, be that in pedigreed populations (e.g., Arbelbide and Bernardo 2006; Breseghello and Sorrells 2006; Kraakman et al. 2004; Parisseaux and Bernardo 2004) or in germplasm collections (Thornsberry et al. 2001). Association mapping models lend themselves readily to modeling a QTL by genetic background interaction because they often already include both main marker and genetic background effect terms (Kennedy et al. 1992; Yu et al. 2006). The genetic background term is necessary to account for heterogeneous genetic relationships between observed individuals in the experiment. That is, some individuals may be more closely related than average by virtue of belonging to the same sub-population, or by belonging to the same (possibly extended) family. The genetic background term therefore prevents error residuals from being correlated, which otherwise would invalidate the test (Jannink et al. 2001). Modeling the QTL by genetic background term therefore amounts to adding an interaction between two terms that are already in the model.

Depending on what type of genetic background term is used, the QTL by genetic background interaction will have different interpretations. The simplest accounting for genetic background entails a term that classifies observations according to what subpopulation the individual that produced them belongs. Note that an individual need not belong exclusively to a single subpopulation. The classification of individuals works equally well if classification variables are continuous and sum to one, as will happen if only the probability that an individual belongs to a particular subpopulation can be inferred (Pritchard et al. 2000; Yu et al. 2006). The interaction between a marker and such a classification indicates that the associated effect differs between subpopulations. The difference in associated effect can have at least two non-mutually-exclusive causes. First, epistasis may be involved. That is, the associated causal polymorphism interacts with other loci that modulate the phenotypic effects of each variant, and allele frequencies at these other loci differ across the subpopulations. Second, subpopulations may

differ in the extent and sign of linkage disequilibrium between a marker locus and causal polymorphism. Hill and Robertson (1968) studied disequilibrium between neutral loci in finite populations and showed that the variance of the linkage disequilibrium coefficient D (defined in Lynch and Walsh 1998, p. 94) across replicate lines drawn from a population in linkage equilibrium could be "of an order of magnitude similar to that of the variance of gene frequencies after some generations of inbreeding." Thus, if drift played an important role in establishing the association between marker and QTL, heterogeneity of association across subpopulations will be the norm. In this case, rather than modeling a main marker effect and a subpopulation interaction effect, it may make more sense to nest the marker effect within subpopulations.

In a somewhat more fine-grained fashion, the genetic background can be modeled as a random effect for each individual, with the variance-covariance matrix of the vector of effects determined by identity by descent (IBD) probabilities among individuals (Lynch and Walsh 1998, p. 755). Usually this model presupposes that all individuals belong to the same subpopulation such that questions of association heterogeneity do not come into play. In this case, the interaction between a given marker allele and the genetic background also becomes a random effect. The interpretation of the effect is that the allele's influence on the phenotype should display resemblance between relatives: the contribution of the allele to the phenotypes of full-sibs will be similar but it's contribution to the phenotypes of two unrelated individuals may be dissimilar.

In this study, I review and develop theory presented in (Jannink 2007) that relates the variance of the marker by genetic background interaction variance back to the standard additive-by-additive epistatic variance (Lynch and Walsh 1998). I then discuss a statistical model to fit this interaction effect in the context of a mixed model of the kind proposed by Kennedy et al. (1992) where the relationship between individuals is accounted for using the pairwise IBD matrix. Simulations test the ability of the model to detect epistasis affecting a previously identified locus and to improve the prediction of the genotypic value of progeny for which marker data but no phenotypic data exists. Previous research with this model indicated that the interaction effect of the genetic background was poorly estimated in individuals that were heterozygous at the focal locus (Jannink 2007). Thus, in simulations performed for this study, I particularly look at how well epistatic effects are explained under two levels of inbreeding.

Theory

To begin with, consider a two-locus model with a focal locus Q and a background locus B. The only effects that need concern us here are the main effects of Q (denoted by α) and its interaction effects with B (denoted by ε). Thus the genotypic value G becomes

$$G = \mu + \alpha_i + \alpha_j + \varepsilon_{ik} + \varepsilon_{il} + \varepsilon_{jk} + \varepsilon_{jl}$$

where i, j, k, and l subscripts indicate the maternal and paternal alleles for loci Q and B, respectively. The α and ε parameters are defined such that their expectations and covariances are zero. With this definition, the additive-by-additive epistatic variance generated by the QB locus pair [denoted σ^2_{AA}(QB)] is equal to four times the variance of a randomly sampled ε parameter in a population in Hardy-Weinberg and linkage equilibrium.

Using this model, we can take into account the genotype at locus B in determining the genotypic value conferred to an individual x by receiving a Q_1 allele at the Q locus: it will not just be α_1 but $\alpha_1 + \varepsilon_{1k} + \varepsilon_{1l}$, which we denote $\alpha_1 + \tau_{1x}$. The parameter τ_{1x} therefore represents the deviation from the main effect of the Q_1 allele conferred to x by virtue of its genetic background at locus B. A parameter τ_{2x} can be similarly defined. These τ parameters are the marker by genetic background interaction effects. To treat these as random effects, we need to determine the covariance of the deviations present in different individuals. Take the covariance of τ_{1x} and τ_{1y}.

$$\mathrm{cov}(\tau_{1x}, \tau_{1y}) = \mathrm{cov}(\varepsilon_{1k} + \varepsilon_{1l}, \varepsilon_{1k'} + \varepsilon_{1l'})$$

Where k' and l' are the alleles at the B locus in y. This covariance decomposes into a sum of four covariances like cov(ε_{1k}, $\varepsilon_{1k'}$). If alleles B_k and $B_{k'}$ are not identical by descent, this covariance is zero. If alleles B_k and $B_{k'}$ *are* identical by descent, that covariance is equal to the variance of a randomly sampled value among all ε_{1k} [denoted var($\varepsilon_{1\bullet}$)]. Consequently,

$$\mathrm{cov}(\tau_{1x}, \tau_{1y}) = 4\mathrm{var}(\varepsilon_{1\bullet})\theta_{xy}$$

where θ_{xy} is the coefficient of coancestry between x and y.

Across all observations, the vector of interaction effects between the Q_1 allele and the genetic background can be modeled with a multivariate normal distribution

$$\boldsymbol{\tau}_1 \sim N(0, 2\mathrm{var}(\varepsilon_{1\bullet})\mathbf{A})$$

where $\mathbf{A}$ is the additive relationship matrix (each element of $\mathbf{A}$ describes the additive relationship between the individual x represented across the row and the individual y represented down the column, $a_{xy} = 2\theta_{xy}$). The matrix $\mathbf{A}$ can be calculated from the pedigree (Lynch and Walsh 1998) or from a sufficient set of DNA markers (Weir et al. 2006). Note that an individual need not carry any Q_1 alleles for it to have a τ_1 effect associated with it. The interpretation of the τ_1 effect in the absence of a Q_1 allele is simply that that is the deviation from the main effect of the Q_1 allele that would obtain if the allele were present in the individual (say, if it were introduced by site-specific mutagenesis). The distribution of the $\boldsymbol{\tau}_2$ vector can be similarly modeled, replacing $\mathrm{var}(\varepsilon_{1\bullet})$ with $\mathrm{var}(\varepsilon_{2\bullet})$.

To relate the variances $\mathrm{var}(\varepsilon_{1\bullet})$ and $\mathrm{var}(\varepsilon_{2\bullet})$ back to the more well-known additive-by-additive epistasis, $\sigma^2_{AA}(\mathrm{QB})$, we need to relate the variance of a randomly sample interaction effect ε, $\mathrm{var}(\varepsilon)$ to them. Two equations that derive from the genetic model are required. First, for any given background allele, B_k, the interaction effects have zero expectation:

$$p_1\varepsilon_{1k} + p_2\varepsilon_{2k} = 0$$

where p_1 and p_2 are the frequencies of the Q_1 and Q_2 alleles ($p_1 + p_2 = 1$). Second, $\mathrm{var}(\varepsilon)$ is a weighted average of interaction effects with the Q_1 allele and interaction effects with the Q_2 allele:

$$\mathrm{var}(\varepsilon) = p_1\mathrm{var}(\varepsilon_{1\bullet}) + p_2\mathrm{var}(\varepsilon_{2\bullet})$$

Using the first equation we find

$$p_1^2\mathrm{var}(\varepsilon_{1\bullet}) = p_2^2\mathrm{var}(\varepsilon_{2\bullet})$$

$$\frac{p_1^2}{p_2}\mathrm{var}(\varepsilon_{1\bullet}) = p_2\mathrm{var}(\varepsilon_{2\bullet})$$

And substituting this into the second equation gives

$$\mathrm{var}(\varepsilon) = \frac{p_1}{p_2}\mathrm{var}(\varepsilon_{1\bullet})$$

Or, through a similar development,

$$\mathrm{var}(\varepsilon) = \frac{p_2}{p_1}\mathrm{var}(\varepsilon_{2\bullet})$$

The distribution of the vector of interactions with Q_1 can therefore be rewritten as

$$\boldsymbol{\tau}_1 \sim N[0, (1/2)\frac{p_2}{p_1}\sigma^2_{AA}(\mathrm{QB})\mathbf{A}]$$

So far, only the interaction between one background locus and the focal locus Q has been treated. To extend this treatment to interactions with *many* background loci, we must make the simplifying assumptions that all interacting background loci are in linkage equilibrium and that they only interact with the focal locus, not with each other. In that case, the overall interaction between the focal locus and the genetic background is simply the sum of its interactions with each individual locus: $\sigma^2_{AA}(\mathrm{Q}) = \sum\sigma^2_{AA}(\mathrm{QB})$. In what follows, the interaction deviation τ_1 is redefined to denote the overall interaction rather than the interaction with a specific locus B.

To cast this genetic theory into a linear model whose parameters can be estimated, consider an observation on individual i

$$y_i = \mu + x_{1i}(\alpha_1 + \tau_{1i}) + x_{2i}(\alpha_2 + \tau_{2i}) + u_i + e_i$$

where μ is the population mean, x_{1i} and x_{2i} are, respectively, the number of Q_1 and Q_2 alleles carried by individual i ($x_{1i} + x_{2i} = 2$), u_i is a polygenic effect that accounts for additive genetic effects not absorbed by the focal locus, e_i is an error residual, and the α and τ terms have already been defined. Note that the $(\alpha + \tau)$ terms have zero expectation:

$$p_1(\alpha_1 + \tau_{1i}) + p_2(\alpha_2 + \tau_{2i}) = 0$$

$$(\alpha_2 + \tau_{2i}) = -(p_1/p_2)(\alpha_1 + \tau_{1i})$$

To make the linear model estimable, replace x_{2i} by $2 - x_{1i}$ and $(\alpha_2 + \tau_{2i})$ with the above expression to get

$$y_i = \mu + [-2(p_1/p_2) + (1/p_2)x_{1i}](\alpha_1 + \tau_{1i}) + u_i + e_i$$

This model can be written in matrix notation as

$$\boldsymbol{y} = \boldsymbol{\mu} + \mathbf{X_1}\alpha_1 + \mathrm{diag}(\mathbf{X_1})\boldsymbol{\tau_1} + \boldsymbol{u} + \boldsymbol{e}$$

Epistatic Model

where $\mathbf{X_1}$ is a vector of the $[-2(p_1/p_2) + (1/p_2)x_{1i}]$ values. Thus, the vector $\boldsymbol{\tau_1}$ has an incidence matrix, which is similar to the incidence of the main associated locus effect and a variance-covariance matrix that is proportional to that of the vector of polygenic effects $\boldsymbol{u}$. Considering all fixed and random effects then, the distribution of the observations according to this model is

$$\boldsymbol{y} \sim N[\boldsymbol{\mu} + \mathbf{X_1}\alpha_1, \mathrm{diag}(\mathbf{X_1})\mathbf{A}\mathrm{diag}(\mathbf{X_1})(1/2)\frac{p_2}{p_1}\sigma^2_{AA}(Q) + \mathbf{A}\sigma^2_u + \mathbf{I}\sigma^2_e]$$

A contrasting reduced statistical model that assumes additive gene action can be defined as nested within the Epistatic model:

$$\boldsymbol{y} = \boldsymbol{\mu} + \mathbf{X_1}\alpha_1 + \boldsymbol{u} + \boldsymbol{e} \quad \text{Additive Model}$$

Simulations

A QTL was marked perfectly such that marker and QTL were in complete LD with no recombination. All other loci were simulated in linkage equilibrium with each other and with the QTL, as follows. The marked QTL interacted with 9 other independent loci in a compound epistatic network (Cooper et al. 2002) with each interaction generating a variance of S under random mating such that the total epistatic variance was $9S$. Thus the 9 other loci create a polygenic background with which the focal QTL interacts. The expected marginal effect of the QTL was zero. Ten other independent loci segregated, each generating an additive variance of 4 under random mating. These 10 loci, in turn, generate a polygenic background with which the QTL does not interact. A normal deviate with a variance of 40 was added to the genetic value to obtain a phenotypic value. The genetic model assumed that all loci were biallelic with frequencies of 0.5. Small stochastic deviations from the expected allele and genotype frequencies in the simulation, however, could generate deviations in the epistatic variance actually simulated. To avoid these deviations, once the pedigree and the allelic states of individuals were simulated, allele frequencies were calculated and epistatic effects adjusted to obtain the desired genetic variances.

Pedigrees with four generations were simulated using this genetic model. Simulations were performed including a factorial of three parameters, each with two levels: The value of S was either 2.2 or 0.55, such that the epistatic variance was 20 or 5 (leading to total phenotypic variances of 100 or 85 under random mating); Each generation of the pedigree contained either 80 or 160 individuals; Individuals were either phenotyped and mated at the S_0 generation or the S_4 generation (where the S_4 was obtained from the S_0 by four generations of self-fertilization). The S_4 generation of inbreeding was chosen because it is often at this level of inbreeding that phenotypic evaluations begin in self-pollinated crops. Individuals of the first three generations were considered to have phenotype and marker data while the fourth generation had only marker data. Thus, the population size that contributed information to the estimation of parameters was either 240 or 480.

Simulated phenotypes were analyzed with a linear mixed model using the Epistatic and Additive models given above. The pedigrees, genotypic, and phenotypic values were simulated, and the $\mathbf{A}$ and $\mathrm{diag}(\mathbf{X_1})\mathbf{A}\mathrm{diag}(\mathbf{X_1})$ matrices needed to fit the model were calculated using a program written in C++. These matrices were then imported into SAS (SAS Institute, Inc., Cary, NC, USA) and the mixed model analysis was performed in proc mixed. Detection power was determined as the fraction of simulations out of 500 for which the QTL by genetic background interaction variance, $\sigma^2_{AA}(\mathrm{Q})$, was found to be significantly greater than zero with a type I error rate of 0.05. Out of the 4,000 analyses performed, 98.9% converged successfully. The others were discarded. Genotypic values of the individuals from the fourth generation were predicted using, for individual i,

$$\hat{y}_i = \hat{\mu} + X_{1i}(\hat{\alpha}_1 + \hat{\tau}_{1i}) + \hat{u}_i$$

and

$$\hat{y}_i = \hat{\mu} + X_{1i}\hat{\alpha}_1 + \hat{u}_i$$

for the Epistatic and Additive models, respectively, where the estimates of τ_{1i}, and u_i derived from the pedigree of i, and the QTL genotype information of i provided X_{1i}. Predicted genotypic values from the two models were then correlated with the true genotypic values known from the simulation.

Results and discussion

The power to detect QTL by genetic background follows what might be expected under random mating: it increases when the size of the epistatic interaction increases and when the population used for detection increases (Table 1). More interestingly, power to detect the interaction greatly increased when the individuals evaluated were inbred such that when the epistatic variance was 20, detection was almost certain. Under the smaller epistatic variance of 5, power was much lower, but still markedly improved under inbreeding relative to random mating. Two hypotheses can be proposed for the increase in power. First, the variance generated by additive × additive effects increases under inbreeding relative to random mating, much as for simple additive effects. Thus, additive × additive effects that generate a variance of V_{AA} in a randomly mating reference population will generate a variance of $4V_{AA}$ in a completely inbred population. Second, note that the value of the incidence matrix parameter $X_{1i} = [-2(p_1/p_2) + (1/p_2)x_{1i}]$ is zero when the individual i is heterozygous ($x_{1i} = 1$) and the marker has intermediate allele frequency ($p_1 = p_2 = 0.5$). Consequently, the phenotype of an individual that is heterozygous at the marker does not contribute information to the estimation of the τ_1 effects, and having more homozygous individuals at the marker should increase the power of the analysis (Jannink 2007).

The analysis appears to estimate the effect of the QTL by genetic background interaction in a relatively unbiased way, regardless of the epistatic variance or the level of inbreeding (Table 1). The lack of bias is striking, considering the point made above that as inbreeding increases, the phenotypic variance generated by additive × additive effects increases. The construction of the coefficients in the incidence vector $\mathbf{X_1}$ ensures unbiased estimation of the variance generated in the population of reference used here, namely a randomly-mating population. When the QTL by background interaction term was included in the model, the additive polygenic and the error variance were also estimated without bias (data not shown). However, in the absence of this term, the additive × additive effects inflated both of those variances (data not shown). Finally, in addition to substantially increasing the power to detect epistatic variance affecting a marked QTL, inbreeding also greatly reduced the error with which the variance was estimated (Table 1). The observed mean squared errors on the epistatic variances decreased by close to 50% or more.

Correlations between the simulated (true) genotypic values of individuals in the fourth generation of the pedigree and the genotypic values predicted by the Epistatic and Additive models provide insight into how epistatic variance affects our ability to predict progeny values (Table 2). Under an additive genetic model and random mating, the regression of progeny value on parent means is equal to the fraction of the phenotypic variance that is additive, that is, the narrow-sense heritability. For the S_0 inbreeding level, the Additive statistical model consistently outperforms this expectation (Table 2). The reasons are that, also under random mating, one fourth of the additive × additive epistatic variance contributes to parent-offspring resemblance. In addition, given that the full pedigree was used for predictions, information from relatives other than parents will have contributed. For the S_4 inbreeding

Table 1 Power to detect QTL × genetic background interaction, mean estimate of the epistatic variance, and mean squared error (MSE) of the estimate over 500 simulations as a function of the size of the population analysed, the epistatic variance simulated, and the inbreeding level of the individuals evaluated

Epistatic variance	Population analysed	Inbreeding level	Detection power	Mean estimate	MSE
20	240	S_0	0.65	20.4	86.0
20	240	S_4	0.94	20.9	42.7
20	480	S_0	0.92	20.0	40.4
20	480	S_4	1.00	20.0	24.6
5	240	S_0	0.08	5.3	47.5
5	240	S_4	0.21	5.1	13.8
5	480	S_0	0.16	5.0	21.9
5	480	S_4	0.50	5.1	7.5

level, predictions of this sort would require theoretical developments that are beyond the scope of this article. Suffice it to say that the predictive ability of the Additive model does not follow an obvious pattern relative to the fraction of the phenotypic variance generated by additive effects: the correlation is higher than the additive variance fraction under high epistasis but lower than additive variance fraction under low epistasis. Still, not surprisingly, increasing epistatic variance from 5 to 20 decreased the predictive ability of the additive model: across other parameter values, correlations averaged about 0.53 versus 0.47 under low and high epistasis respectively. Epistasis obviously affects genotypic values in ways that the Additive model cannot account for and that, therefore, increase error.

Less expected was the observation that increasing epistatic variance did not improve the ability of the Epistatic model to predict progeny genotypic values. Apparently the Epistatic model was not able to capture and account for all of the variance in genotypic values generated by the interacting loci. Under the S_0 inbreeding level, the total genotypic variance (additive + epistatic) increased from 45/85 = 53% to 60/100 = 60% of the phenotypic variance under low versus high epistatic variance, respectively. But this increase did not translate into an increase in the correlation observed between simulated and predicted progeny genotypic values. Similarly, under the S_4 inbreeding level, the total genotypic variance increased from 70% to 78% of the phenotypic variance under low versus high epistatic variance without improving the ability of the Epistatic model to predict progeny genotypic values (Table 2). Perhaps rather than seeking to understand the behavior of these models relative to theoretical predictions, we may better simply compare their relative performance. Here the results are fairly straightforward: greater epistatic variance improved the Epistatic model relative to the Additive model; Higher levels of inbreeding also improved the Epistatic model relative to the Additive model, consistent with the fact that the variance generated by additive × additive effects increases more rapidly than that of additive effects. Finally, the different population sizes had almost no perceptible effect on the mean correlations observed for the two models. In terms of the probability of the Epistatic model outperforming the Additive model in any given simulation, however, larger populations benefited the Epistatic model over the additive model (last column of Table 2). The advantage conferred to the Epistatic model from more observations presumably comes from the fact that the model requires the estimation of an additional variance component. That component is estimated more accurately when larger populations are available (Table 1). Extrapolating from Table 2, we can conjecture that the minimal epistatic variance required for the Epistatic model to outperform the Additive model is about 5% under random mating, and about 2% under inbreeding for the population sizes envisioned here. These are fairly substantial variances to be associated with a single marked locus.

Finally, it is worth contrasting the QTL by genetic background interaction model presented here with

Table 2 Mean correlation between the simulated genotypic value of a fourth generation individual and its predicted genotypic value using the Epistatic or Additive models, and the frequency with which the correlation was higher for the Epistatic model than for the Additive model (Probability Epi > Add). Individuals for which the correlation was evaluated had DNA marker data but no phenotypic record

Epistatic variance	Population analysed	Inbreeding level	Additive correlation	Epistatic correlation	Additive fraction	Epistatic fraction	Probability Epi > Add
20	240	S_0	0.46	0.49	0.40	0.20	0.80
20	240	S_4	0.46	0.55	0.40	0.38	0.93
20	480	S_0	0.47	0.50	0.40	0.20	0.86
20	480	S_4	0.47	0.56	0.40	0.38	1.00
5	240	S_0	0.50	0.50	0.47	0.06	0.51
5	240	S_4	0.54	0.55	0.57	0.13	0.69
5	480	S_0	0.51	0.51	0.47	0.06	0.56
5	480	S_4	0.54	0.56	0.57	0.13	0.79

others that have been published in the literature. First are the QTL by genetic background models that can be used when a diallel of crosses has been performed (Blanc et al. 2006; Charcosset et al. 1994; Jannink and Jansen 2001). In this case, the "genetic background" is modeled in a very coarse way: each cross or population constitutes a background. Differences between individuals within a population cannot be accounted for with these methods. This coarseness may account in part for the partial failure of the methods in a recent case (Blanc et al. 2006). In this case, the QTL by genetic background test did not identify loci that were not already identified in tests between pairs of loci. Blanc et al. (2006) noted the possibility of "a repartition of alleles among parents so that several digenic epistatic effects cancel out each other and result in no significant QTL by genetic background interactions." The canceling out of digenic epistasis of this type would occur because the alleles at interacting loci are averaged over the genetic background considered in those models, that is, over the population as a whole. Methods that can account for genetic background on a finer scale, that is, individual by individual, might have a better chance at identifying such interactions. Indeed, in the case simulated here, the locus that interacted with the background was part of nine digenic interactions. Had the genetic background been modeled in the coarse way of the diallel cross approach, those interactions might indeed have canceled each other out.

At an even finer scale, Boer et al. (2002) envisioned one-dimensional genome scans for epistasis in which all pairwise interactions between the focal locus and all other markers were tested simultaneously. Because this approach requires many factors to be included in a single model, ridge regression was adopted to penalize large effect estimates (Boer et al. 2002; Whittaker et al. 2000). The Boer et al. (2002) method is in some sense a hybrid between the QTL by genetic background approach and standard digenic interaction approaches. Because it examines one focal locus at a time, it identifies loci that interact with background. At the same time it should identify those regions of the genome that interact with the focal locus, something that QTL by genetic background approaches typically do not do. Balancing these advantages is the fact that the Boer et al. (2002) method requires a set of markers in linkage disequilibrium with their surrounding genome, such that reasonable genome coverage is obtained. In the absence of such coverage, interactions may be missed. Linkage studies typically achieve this coverage (and Boer et al. 2002 examined the method in that context) but association studies may not, particularly when they focus on candidate loci. The method presented here does not require such coverage.

In conclusion, therefore, the method presented here presents a hopefully useful middle ground to identify loci that interact with the genetic background even when their interactors cannot be located. The most striking result of this contribution is the increased accuracy of epistatic variance estimation and progeny genotypic value prediction that occur when the method is applied to inbred individuals. This should make the method well-suited to self-pollinating crops such as wheat, barley, oat and soybean.

Acknowledgments This research was supported by USDA-NRI grant number 2003-35300-13202. I thank Fred van Eeuwijk, Jerko Gunjaca and other organizers of the 2006 EUCARPIA Biometrics Section Meeting for an excellent conference.

References

Arbelbide M, Bernardo R (2006) Mixed-model QTL mapping for kernel hardness and dough strength in bread wheat. Theor Appl Genet 112:885–890

Blanc G, Charcosset A, Mangin B, Gallais A, Moreau L (2006) Connected populations for detecting quantitative trait loci and testing for epistasis: an application in maize. Theor Appl Genet 113:206–224

Boer MP, ter Braak CJF, Jansen RC (2002) A penalized likelihood method for mapping epistatic quantitative trait loci with one-dimensional genome searches. Genetics 162:951–960

Breseghello F, Sorrells ME (2006) Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. Genetics 172:1165–1177

Bryant EH, Meffert LM (1993) The effect of serial founder-flush cycles on quantitative genetic variation in the housefly. Heredity 70:122–129

Carlborg O, Jacobsson L, Ahgren P, Siegel P, Andersson L (2006) Epistasis and the release of genetic variation during long-term selection. Nat Genet 38:418–420

Charcosset A, Causse M, Moreau L, Gallais A (1994) Investigation into the effect of genetic background on QTL expression using three connected maize recombinant inbred lines (RIL) populations. In: v. Ooijen JW, Jansen J (eds) Biometrics in plant breeding: applications of molecular markers. CPRO-DLO, Wageningen, The Netherlands, pp 75–84

Cheverud JM, Routman EJ (1996) Epistasis as a source of increased additive genetic variance at population bottlenecks. Evolution 50:1042–1051

Cheverud JM, Vaughn TT, Pletscher S, King-Ellison K, Bailiff J, Adams E, Erickson C, Bonislawski A (1999) Epistasis and the evolution of additive genetic variance in populations that pass through a bottleneck. Evolution 53:1009–1018

Cooper M, Podlich DW, Micallef KP, Smith OS, Jensen NM, Chapman SC (2002) Complexity, quantitative traits and plant breeding: A role for simulation modeling in the genetic improvement of crops. In: Kang MS (ed) Quantitative genetics, genomics and plant breeding. CABI Publishing, Wallingford, U.K., pp 143–166

Frey KJ, Holland JB (1999) Nine cycles of recurrent selection for increased groat-oil content in oat. Crop Sci 39:1636–1641

Goodnight CJ (1987) On the effect of founder events on epistatic genetic variance. Evolution 41:80–91

Goodnight CJ (2004) Gene interaction and Selection. In: Lamkey KR (ed) Long-term selection: maize. plant breeding reviews, Vol. 24. John Wiley & Sons, Inc., Hoboken, New Jersey, pp 269–292

Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. Theor Applied Genet 38:226–231

Jannink J-L (2003) Selection dynamics and limits under additive-by-additive epistatic gene action. Crop Sci 43:489–497

Jannink J-L (2007) Identifying quantitative trait locus by genetic background interactions in association studies. Genetics 176:553–561

Jannink J-L, Jansen RC (2001) Mapping epistatic QTL with one-dimensional genome searches. Genetics 157:445–454

Jannink J-L, Bink MCAM, Jansen RC (2001) Using complex plant pedigrees to map valuable genes. Trends Plant Sci 6:337–342

Kennedy BW, Quinton M, van Arendonk JAM (1992) Estimation of effects of single genes on quantitative traits. J Anim Sci 70:2000–2012

Kianian SF, Egli MA, Phillips RL, Rines HW, Somers DA, Gengenbach BG, Webster FH, Livingston SM, Groh S, LS OD, Sorrells ME, Wesenberg DM, Stuthman DD, Fulcher RG (1999) Association of a major groat oil content QTL and an acetyl-CoA carboxylase gene in oat. Theor Appl Genet 98:884–894

Kraakman ATW, Niks RE, Van den Berg PMMM, Stam P, Van Eeuwijk FA (2004) Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. Genetics 168:435–446

Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits. Sinauer Associates, Sunderland, MA

Parisseaux B, Bernardo R (2004) In silico mapping of quantitative trait loci in maize. Theor Appl Genet 109:508–514

Podlich DW, Cooper M (1998) QU-GENE: a platform for quantitative analysis of genetic models. Bioinformatics 14:632–653

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES (2001) Dwarf8 polymorphisms associate with variation in flowering time. Nat Genet 28:286–289

Weir BS, Anderson AD, Hepler AB (2006) Genetic relatedness analysis: modern data and new challenges. Nat Rev Genet 7:771–780

Whittaker JC, Thompson R, Denham MC (2000) Marker-assisted selection using ridge regression. Genet Res 75:249–252

Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES IV (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203–208